LAWRENCE LIVERMORE NATIONAL LABORATORY

LLNL-JRNL-484206

# Optical Spectra Analysis for Breast Cancer Diagnostics

S. A. Belkov, G. G. Kochemasov, T. E. Lyubynskaya, N. V. Maslov, A. S. Nuzhny, L. B. Da Silva, A. Rubenchik

May 17, 2011

Applied Physics B

**Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

# Optical Spectra Analysis for Breast Cancer Diagnostics

**S.A.Belkov[1], G.G.Kochemasov[1], T.E.Lyubynskaya[1], N.V.Maslov[1], A.S.Nuzhny[2], L.B. Da Silva[3], and A.Rubenchik[4]**

**Abstract:** Minimally invasive probe and optical biopsy system based on optical spectra recording and analysis seem to be a promising tool for early diagnostics of breast cancer. Light scattering and absorption spectra are generated continuously as far as the needle-like probe with one emitting and several collecting optical fibers penetrates through the tissues towards to the suspicious area. That allows analyzing not only the state of local site, but also the structure of tissues along the needle trace. The suggested method has the advantages of automated on-line diagnosing and minimal tissue destruction and in parallel with the conventional diagnostic procedures provides the ground for decision-making.
About 200 medical trials were completed in Nizhny Novgorod Regional Oncology Centre, Russia. Independent diagnoses were the results of fine biopsy and histology.
Application of wavelet expansion and clasterization techniques for spectra analysis revealed several main spectral types for malignant and benign tumors. Automatic classification algorithm demonstrated specificity ~ 90% and sensitivity ~91%.
Large amount of information, fuzziness in criteria and data noisiness make neural networks to be an attractive analytic tool. The model based on three-layer perceptron was tested over the sample of 29 'cancer' and 29 'non-cancer' cases and demonstrated total separation.

## 1 Introduction

Breast cancer is known to be the most widespread female oncology disease and reason of female mortality in the world. But being discovered at early stage it may be successfully treated by combination of surgery, chemotherapy and radiation. All diagnostic method currently used in clinical practice have some imperfections and are unable to provide the indexes of sensitivity and specificity high enough for reliable diagnosing. That is why novel techniques for breast cancer diagnostics at early stages are actively developed all over the world.
In the international science the diagnostic ability of optical spectra of biological tissues is widely studied. As it was shown in [1-3] the shapes of the spectral curves have specific absorption bands, which are deformed when the disease progresses, and that allows tissue differentiation on types and conditions. Application of a contact probe for surface cancer diagnostics was demonstrated in [4, 5]. That probe has one emitting and one collecting light fibers measuring light scattering spectrum in the range of 350-700 nm.
Minimally invasive probe and the diagnostic system developed by Biotilligent (USA), Biofil (Russia) and Russian Federal Nuclear Centre – VNIIEF are based on optical radiation spectra recording and analysis and seem to be a promising tool for early diagnostics of breast cancer [6]. In this system optical scattering spectra are generated continuously as far as the needle-like probe containing one emitting and several collecting optical fibers penetrates through the tissues moving to the suspicious area. That allows analyzing not only the state of local site, but also the structure of tissues along the needle trace. The suggested method has the advantages of automated on-line diagnosing and minimal tissue destruction and in parallel with the traditional diagnostic procedures provides the ground for decision-making.

## 2 Technical details

The system consists of a source of white light (Xe lamp), a measurement circuit and a unique probe with disposable needles incorporating optical fibers, and a computer for operation control and data recording. The flow diagram of the diagnostic system is presented in fig.1.

[1] Russian Federal Nuclear Center-VNIIEF, 37 Prospekt Mira, Sarov, Nizhny Novgorod reg., 607200, Russia, e-mail tlyubyn@gmail.com
[2] Nuclear Safety Institute of Russian Academy of Science, 52 Bolshaya Tulskaya, Moscow, 115191, Russia
[3] BioTelligent Inc., 6248 Preston Ave., Livermore, CA 94551, USA
[4] LLNL, 7000 East Ave., Livermore, CA 94551, USA

The xenon lamp has continuous spectrum in the range of 370-750 nm. Configuration of the probe and diagnostic system allows obtaining of quantitative characteristics of optical scattering and absorption in various kinds of biological tissues. A photo of the diagnostic system is given in fig. 2.

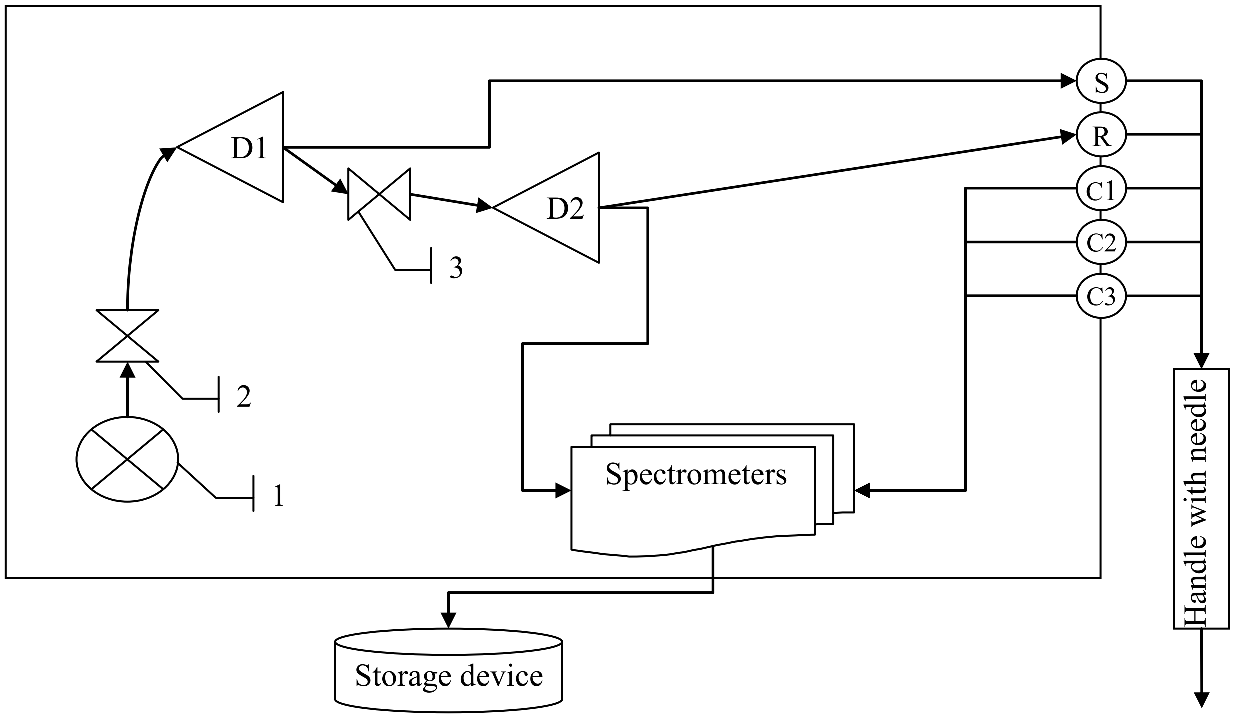



Fig. 1 Flow diagram of the diagnostic system: 1 – xenon lamp; 2 – matching unit; D1, D2 – optical-fiber splitters; 3 - attenuator; S – white light source channel; R – reference channel for source spectrum measurement and calibration; C1, C2, C3 – scattered radiation measuring channels

Fig.3. represents a photo of the optical probe and schematic of fibre location on the needle tip. The needle diameter is 0.8 mm and the length is 50 mm, fiber diameter is 100 μm. One of the fibers inside the needle emits white light, which interacts with biological tissue and then is collected by three other fibers located on the different distances (several hundred microns) from the source. The shape of spectral curves depends on the geometry. Ideally, fibers 2 and 6 being on the same distance from the emitting fibre should give the same results assuming the tissue structure does not change over the probe cross-section. Additional data from the fibre 5 in principle allows reconstruction of local elastic scattering coefficients. To control the depth of needle penetration and define the mechanical parameters of biological tissues the optical probe was provided with position and force sensors.

The measurement system consists of three S2000 fiber optic spectrometers (*Ocean Optics Inc*., USA) with the range from 200 to 1100 nm and resolution ~1.76 nm.

Each PC record contains the data of a microscopic volume of tissue. Macroscopic information about tissue optical properties is obtained by continuous data acquisition during the whole period of probe movement. Recording frequency 100-120 Hz enables spectral measurements each 100 μm along the trajectory of needle movement at the recommended rate 1 cm/sec.

To eliminate the spectral distortions accumulated along the optical path inside the system and on the optical contact between the needle and the handle special calibration techniques were developed. As the needles are disposable calibration is required after each optical biopsy procedure to unify the data acquired with different needles. The spectral response of the needle cleaned from remains of tissue is obtained in the turbid media with the calibrated optical properties. As calibrated media we use 1 or 10% water solution of polystyrene balls calibrated in size (1 μm in diameter). The spectrum of scattering coefficients is calculated using Mie theory.

Optical probing is to be performed after the tumor is discovered by palpation or mammography. The procedure of optical biopsy is similar to that of fine-needle aspiration biopsy (FNA) and may be used every time when fine biopsy is recommended, but can be cheaper and produce the instant result. In addition to breast cancer investigation, which is being conducted and reported here minimally invasive optical probing may be applied to other organs: prostate, thyroid gland and other parenchyma organs.

Fig. 2. Photo of the diagnostic system


Fig. 3. Photo of the optical probe (left) and schematic of fiber location on the needle tip (right).
1 – white light emitting; * - scattered light collecting

## 3 Data acquisition and preliminary processing

To perform clinical studies a test protocol was developed and approved by the Ethical Committee of Nizhny Novgorod State Medical Academy of Ministry of Health of Russian Federation on scientific study with human participation as a subject of investigation. In Nizhny Novgorod Regional Oncology Centre a procedure room was equipped and about 200 tests were completed. Each procedure was supported by video filming and audio recording of the comments of the physician, who carried out the procedure. Independent diagnoses were the results of fine biopsy and sometimes of histology – when excision was required.

Scattered radiation from each collecting fiber is registered over a fixed set of optical wavelengths. After non-informative segments are removed (which are not in the Xe lamp spectral range) the total number of spectral points is 184. An example of starting data is presented in Fig.4. Number of spectra records for an individual patient varies and may be as large as tens of thousands.

The main problem at optical scattering and absorption data interpretation is high level of noise. The reasons are, first, the changes in the properties of the optical contact between the needle and the handle caused by accidental tensions that take place while the needle moves through the breast tissue. That gives random drops in the scattered light intensity.

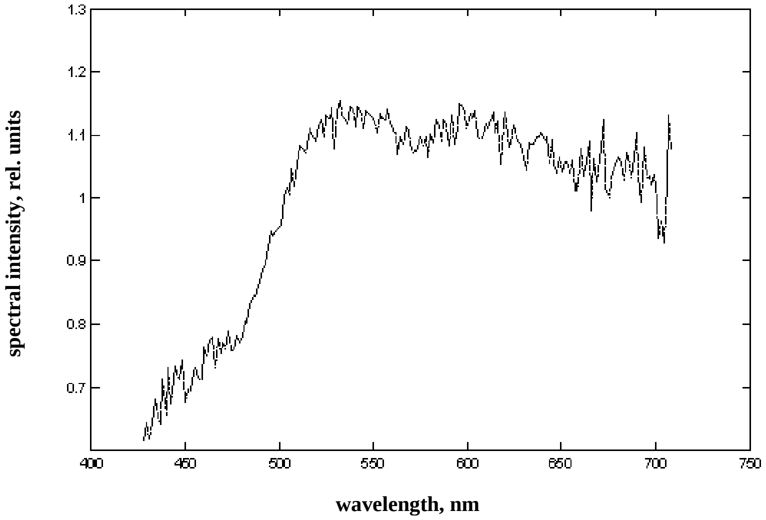



Fig.4. An example of raw data: current optical scattering spectrum

Although performed cross-entropy analysis demonstrated that the signal intensity is quite informative value, in the present study we use normalized spectra, which further simplifies data interpretation. Due to small accumulation time (ca 7 ms) recorded spectra exhibit noticeable detector noise. This noise may be mitigated by averaging the spectral data over the uniform tissue areas. Averaging must be done carefully, because tissue structure may change on the small scale. This is especially true for malignant tumors, which may have inclusions of normal tissue. Examination of normalized and time-averaged spectral curves made it possible to reveal the major spectral distinctions of malignant and benign tumors.

Not all the records are reliable and may be used for the diagnostic purposes. Visual examination discovered that the spectra of skin were close to those of the malignant tissues. So we did not include into consideration the data obtained on the depth less that 3 mm. The data acquired on the back movement were excluded too, because the channel was filled with blood, which distorted real tissue spectra. The position sensor data were used for primary filtration. Fig.5 represents temporal dependence of the needle depth penetration in one of the clinical experiments. X-axis is the number of time points from the beginning of the procedure. Two humps correspond to two insertions done in the different directions. Blue curve represents the initial dataflow. One can see that in the part of time the needle was in the air, moved back or stood still. 'Reliable' points used for diagnosing are marked with red.

Unfortunately, some cases were totally rejected because of either technical or performing reasons. About 150 tests were considered trustworthy enough and made the base for spectral types definition. Only the most reliable of them 29 'cancer' and 29 'non-cancer' cases were selected for neural network development as a prototype of future on-line recognition technique.

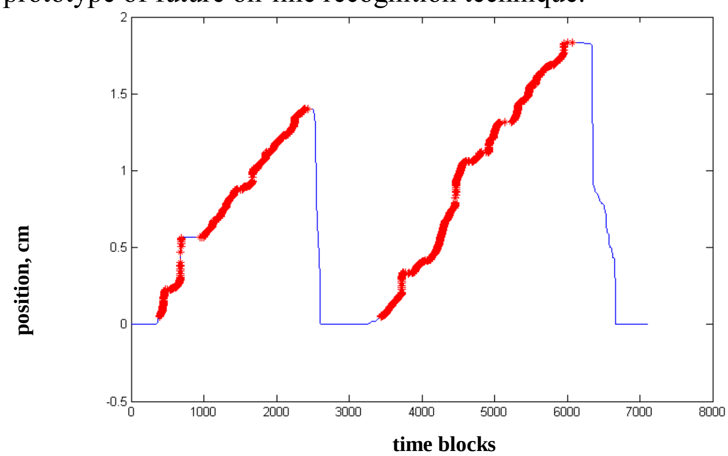


Fig.5. Position sensor readings in one of the clinical trials (blue curve). Red points indicate forward motion of the probe.

## 4 Main spectral types of malignant and benign tumors

Alternatively to time averaging mentioned above, to analyze spectra and reveal ‘malignant’ and ‘benign’ spectral families the method of wavelet expansion and clasterization in the space of wavelet coefficients was used.

Wavelet expansion is widely used for signal processing and filtering, because it enables to get rid off noise and various artifacts in the data such as random surges, gaps, nonlinear distortions, etc. All that may hide essential features in data or pretend to be them and may deteriorate the analysis results dramatically.

Wavelet expansion gives representation of the signal on different scales: approximation of the initial signal of required degree of smoothness (low-frequency component) and set of details, which are the difference between smoothed and initial signals.

Haar wavelet was applied in the present study as the most saving for calculation resources. The allowable wavelet scales are specified by the spectral region and spectral step. For this particular case, 8 levels of expansion can be achieved at the most. Initial clustering was done in 13-dimensional space of the wavelet coefficients at 4$^{th}$ level of expansion.

For malignant tumors, three families of spectral curves were effectively separated. The most representative family contains about 73% of all spectral samples of regions suspected of cancer. The average spectral curves for all of the three identified families (with allowance for standard deviation) are shown in fig. 6. Spectral intensity in figures 6-9 is given in relative units providing unity average.

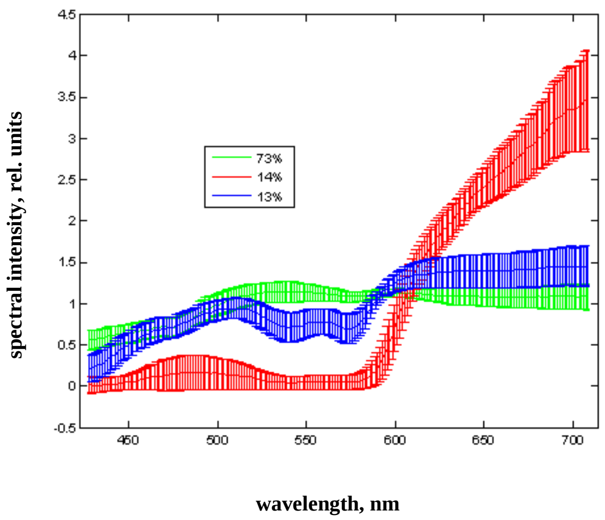



Fig. 6. Three spectral families identified for malignant tumors: curves corresponding to the cluster centres and associated dispersion.

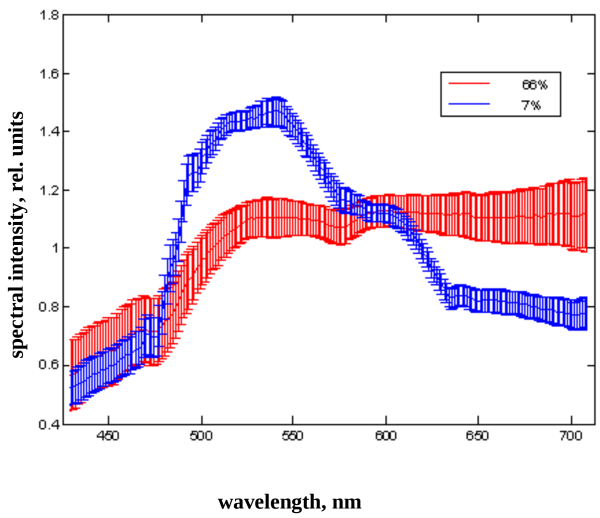
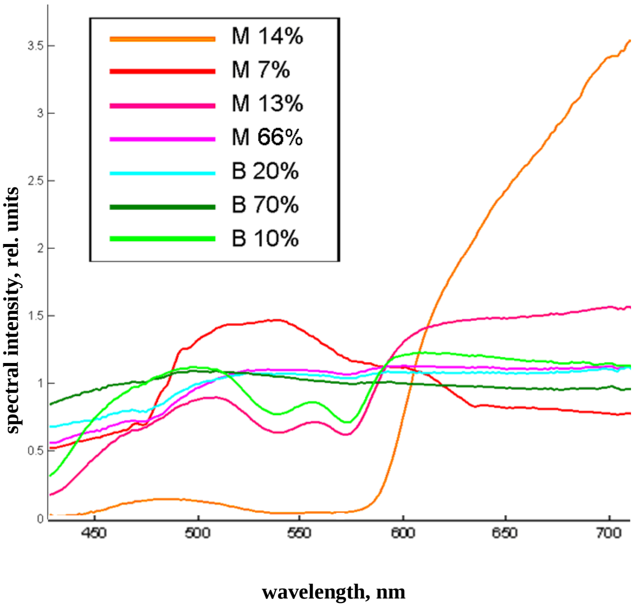



Fig. 7. Two spectral subfamilies composing the most representative family (73%) of malignant tumors: curves corresponding to the cluster centres and associated dispersion

Lower level of expansion enables to see more details. In the 26-dimension space of wavelet coefficients at third level of expansion the most representative family positively separated into two subfamilies. Figure 7 shows average spectral curves for these families with the respective variance.

The similar procedure was also performed to the set of scattering spectra in healthy tissues and benign tumors. As a result the main spectral types were obtained, which are shown in figure 8.

One can see that there are two groups of very close spectral shape: M66 and B20. This caused main difficulties in case separation. In the 1$^{st}$ channel data these groups almost completely overlap within the measuring error. This overlap is smaller in the 3$^{rd}$ cannel data, and the best result that may be obtain from these measurements is the ratio of channel 1 to channel 3 data, which is shown in fig.9 together with experimental errors. Red color corresponds to the malignant cases and green to the benign ones. Although overlap is still quite large, one can see that general slope differs significantly. To reveal some essential details forming the difference between ‘cancer’ and ‘non-cancer’ signals the probability density distributions of their wavelet coefficients on different scales were calculated and only the most divergent were considered. Automatic recognition algorithm was developed basing on two of them. Overlapping of probability density distributions for 'cancer' and 'non-cancer' data are still essential and that put a restriction on the method efficiency.



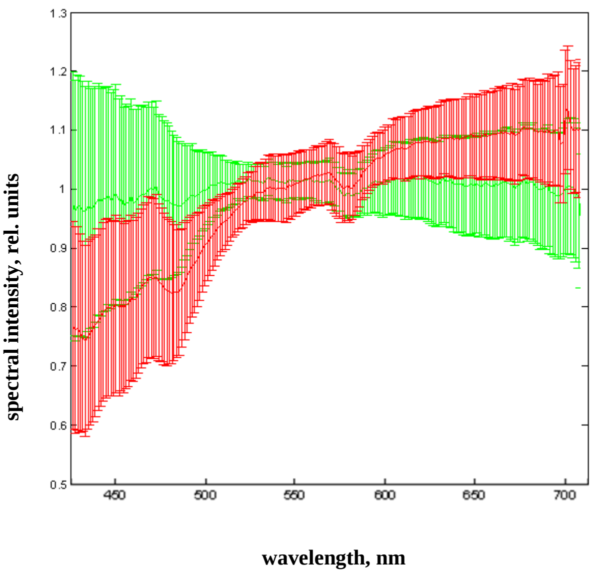




Fig. 8. The set of spectral templates used in automatic detection method. 'M' corresponds to malignant spectra, 'B' – to benign ones. The percentage shows the number of investigated spectra relating to the given type.

Fig. 9. Spectral dependence of the ratio of channel 1 to channel 3 data together with experimental errors. Red color corresponds to the malignant cases and green to the benign ones.

The automatic diagnostic algorithm was developed basing on revealed spectral templates and statistical analysis of wavelet coefficients of the spectra. The automatic diagnoses were compared with those given by physicians. The method demonstrated rather high efficiency parameters: the indexes of sensitivity and specificity were found equal to 90% and 91% correspondingly. We believe that more statistics might improve these parameters.

## 5 Neural network analysis

Large amount of information acquired in each procedure, fuzziness in criteria of 'cancer' family membership and data noisiness make neural networks to be an attractive analytic tool for optical biopsy data. To define the dividing rule between 'cancer' and 'non-cancer' spectral families a three-layer perceptron was applied.

The principles of learning theory [7] require composing a learning sample, which is a set of spectrum examples presenting different tissue types. The learning sample was formed by the experts and only the most reliable data were included. Each spectral envelop was associated with one of the meanings of the binary value $y$: $y = 1$ for 'cancer' spectra and $y = -1$ for 'non-cancer' ones.

To smooth the detector noise it makes sense to average the spectral data over some time window. Besides, spectral envelopes were normalized on average brightness. Random drops in the scattered intensity caused by contact loss in the optical fiber connection points do not permit using in full the current intensity values as a component of analysis. Nevertheless, cross-entropy study demonstrated that average brightness is quite an informative value, so it was included in the set of input parameters. So input data for the perceptron were composed of a set of wavelet expansion coefficients of the spectral curves plus one more parameter – the average brightness of the spectrum.

To build a model for data division the method of learning sample approximation was proposed. Approximation is the problem of reconstruction of the rules of data generation on the ground of finite amount of known data – the inverse problem. As the majority of inverse problems it is ill-posed, i.e. has a set of solutions. Such uncertainty may be eliminated by regularization – restriction of the set of possible solutions. In our study Bayesian regularization method was used [8].

The tissues with high blood content cause some problems for analysis. They may be the result of vessel damage by the needle-like probe. But also there can be a dense vessel net surrounding the malignant tumor. To separate these two cases, the decision was made to compose an additional learning sample

containing ‘cancer’ and ‘non-cancer’ spectra of bloody tissue and teach another perceptron to classify them. The spectra recognized by the first perceptron as ‘cancer’ ones were given on the input of the second perceptron, which gave the final diagnosis. Of course, there was another possible way: to join the learning samples and apply only one but more ‘powerful’ perceptron. But from the mathematical point of view, two sequential models are better than one ‘big’ model, because simultaneous optimization over the large number of parameters has higher probability of getting the model into a local minimum.
To improve the prediction ability independent processing of two geometrically equivalent channels and their voting was included into the model. The voting rule was: if both channels showed ‘cancer’, the decision was: ‘cancer’, otherwise: ‘non-cancer’.
Fig.10 demonstrates temporal dependence of (top-down) position sensor readings, 1st perceptron output, 2nd perceptron output for the first and the second channels and final decision for the ‘cancer’ case. The same set of graphs is given in the Fig.10 for the ‘non-cancer’ case. One can see that fig.10 gives a positive diagnosis and fig.11 gives the negative one, which corresponds to the fine biopsy results.
The model was tested over the sample containing 29 ‘cancer’ and 29 ‘non-cancer’ cases. Final separation was 100%. Nevertheless cancerous tumors are diverse and our limited sampling is far from being all-inclusive. It is rational to assume that if the testing sample increases the sensitivity and specificity will deteriorate. However the developed model has strong potential to generalize a great variety of data and the accuracy of the suggested method may eventually become comparable with that of the most advance methods of breast cancer diagnostics.

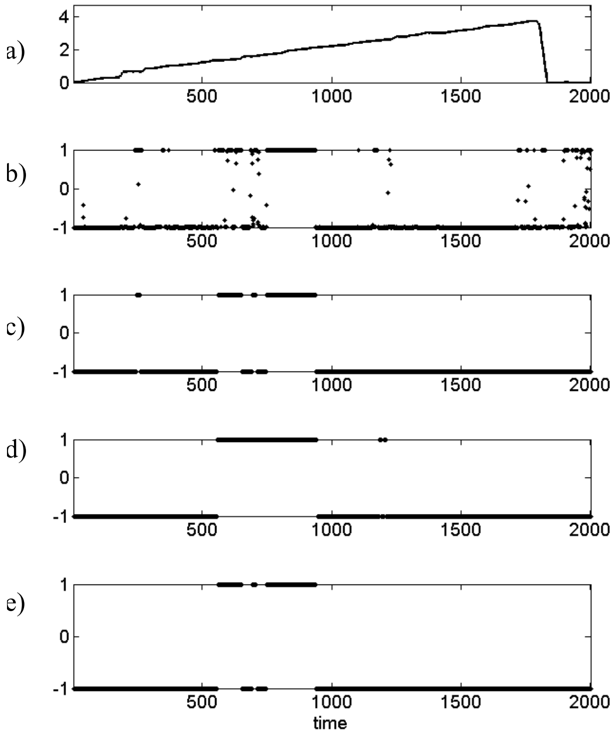



Fig.10. Temporal dependence of position sensor readings (a), 1st perceptron output (b), 2nd perceptron output for the first (c) and the second (d) channels and final decision (e) for the ‘cancer’ case;

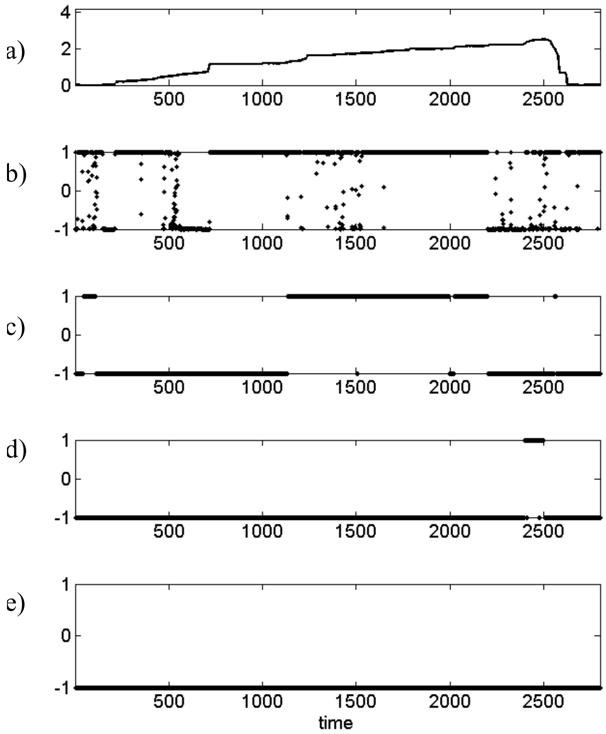



Fig.11. Temporal dependence of position sensor readings (a), 1st perceptron output (b), 2nd perceptron output for the first (c) and the second (d) channels and final decision (e) for the ‘non-cancer’ case.

## 6 Discussion

In recent years optical biopsy methods proved their efficacy for cancer diagnostics, prognostics and treatment management. These methods have advantages of minimal invasiveness, on-line results, low cost, ease of implementation and also they do not require taking out any tissue samples and involving high-skilled pathologists for data interpretation. Fluorescence, Raman and elastic scattering signals are commonly used in oncology studies. The underlying physical principles are described in details in review [9]. In breast cancer studies besides diagnostics [4] optical spectroscopy serves for tumor margin assessment during surgery [10] and lymph node assessment [11]. Basing on optical spectrometry measurements oxygen saturation parameters of the tumor tissue may be reconstructed, which is very valuable from the prognostic point of view [12].

The optical biopsy system under development is minimally invasive and additionally to listed above has a benefit of instant analysis of all tissues surrounding the needle tip along the trace. It was clinically tested in Nizhny Novgorod Regional Oncology Center, Russia. About 200 patients with breast tumors were investigated.
In all cases the optical biopsy procedure was followed by standard fine biopsy procedure and cytological analysis. In some cases tumor tissues were subjected to surgery and histology investigation.
Automatic data processing and analysis algorithm was developed to evaluate acquired spectral data that may extent to tens of thousands spectra per patient.
Expansion of the spectra over Haar wavelets followed by clasterization enabled to reveal several main spectral ‘families’ for malignant and benign tumors. Automatic classification algorithm was developed. The results were compared to the results of medical diagnoses performed by the physicians on the basis of cytology and histology investigation.
The indexes of sensitivity and specificity were found equal to 90% and 91% correspondingly. Earlier we reported these parameters to be 96% and 80% [6]. That algorithm differed significantly from the current approach and by adjusting parameters the sensitivity and specificity can be adjusted. Given the large data set and comparatively small patient sample significant algorithm development and improvement is possible.
The method of artificial neural networks was also applied for data analysis. A three-layer perceptron was used to separate ‘cancer’ and ‘non-cancer’ spectral families. Another perceptron was learnt exceptionally on the spectra with high blood content. The two collecting channels were independently processed and their voting was included into the model. The tests were carried out on the sample of 29 ‘cancer’ and 29 ‘non-cancer’ cases and demonstrated total separation. Although we expect that the sensitivity and specificity may deteriorate with testing sample increase (because of imperfectly performed cases), we believe that the developed model has strong potential to generalize a great variety of data and its accuracy may eventually become comparable with that of the most advance methods of breast cancer diagnostics.
Some technical improvements have been already done basing on the experience obtained in the clinical experiments. The handle of new design was developed with only 4 fibers to minimize the needle diameter. The fibres are connected directly to the front panel for better optical contact. Shock absorber was added to smooth needle movement. New calibration method and special quartz calibration cell were developed to improve reliability of the obtained data.
We believe that the optical biopsy system for the internal tissue test has a great potential for development. First, the procedure of optical probing is similar to fine-needle aspiration biopsy (FNA) and may be used every time when fine biopsy is recommended, but can be cheaper and produce the instant result. In addition to breast cancer investigation that is being conducted optical probing may be applied to other organs: prostate, thyroid gland and other parenchyma organs. Secondly, the physicians collaborated with the project would like to combine in the same device optical biopsy and fine aspiration biopsy, which currently is standard and obligatory procedure at breast oncology investigation in Russia. This is technically feasible and would give the whole body of information in one procedure. For this initial study FNA results and in some cases biopsy serve as the ‘gold standard’, which provides some degree of uncertainty since FNA alone can have false-negative rate exceed 10% [13]. Thirdly, basing on the optical biopsy diagnostic system minimally invasive phototherapy system can be developed usable with and without photosensitizer by replacing Xe lamp with appropriate light source.

## 7 Conclusion

The optical biopsy system underwent first clinical tests. Automatic data processing and analysis algorithm was developed and the indexes of sensitivity and specificity were estimated basing on medical diagnoses: 90% and 91% correspondingly. The method of artificial neural networks was also applied for data analysis as a prototype of further on-line diagnostic algorithm and total separation on the limited dataset was demonstrated.
The study is in the very beginning, but the developed method already seems to be very promising. Comparing with the traditional breast cancer diagnostics it has the advantages of automated on-line diagnosing and minimal tissue destruction. In parallel with the conventional diagnostic procedures optical probing provides the ground for decision-making.

## 8 Acknowledgements

This work was supported by funding from GIPP (Global Initiatives for Proliferation Prevention) Program of US Department of Energy under the contract LLNL-T2-0242-RU and the Project #3075p of International Science and Technology Center. Prepared by LLNL under Contract DE-AC52-07NA27344.

*References*:
[1] J.R. Mourant, J. Boyer, A. Hielscher, I.J. Bigio. Opt. Lett. **21**, 546 (1996)
[2] L.T.Perelman, V. Backman, M.Wallace, et al. Phys. Rev. Lett. **80**, 627 (1998)
[3] J. R. Mourant, M. Canpolat, C. Brocker, et al. J. of Biomedical Opt. **5**, 131 (2000)
[4] I.J.Bigio, S.G.Bown, G.Briggs, et al. J. of Biomedical Optics, **5**, 221 (2000)
[5] G. Zonios, L.T. Perelman, V. Backman, et al. Applied Optics **38**, 6628 (1999)
[6] S.A.Belkov, G.G.Kochemasov, S.M.Kulikov, et al., Proceedings of SPIE **6859**, 20 (2008)
[7] A. Tikhonov, V. Arsenin, Solutions of ill-posed problems (W.H.Winston, Washington, DC 1977)
[8] P. Williams, Neural Computation **7**, 117 (1995)
[9] I.J.Bigio, S.G.Bown, Cancer Biology and Therapy, **3**:3, 259 (2004)
[10] F.T. Nguyen, A.M. Zysk, J.G. Kotynek, et.al., Proceedings of SPIE **6430** A (2007)
[11] W. Luo, F.T. Nguyen, A.M. Zysk, et.al., Tech. Cancer Res. Treat., **4**, 539 (2005)
[12] Brown J.Q., Wilke L.G., Geradts J., et al., Cancer Res., **69**(7), 2919 (2009).
[13] Willis S.L., Ramzy I., Acta Cytologica **39**(5):858, 64 (1995)